# Folksonomy-based Reasoning in Opportunistic Networks

Nishanth Sastry[*]
University of Cambridge
MIT

## ABSTRACT

Disparate algorithms are being designed to decide certain basic questions in opportunistic networks. This position paper describes a nascent idea that aims to provide a single framework to answer such questions. Inspired by the concept of a generic knowledge plane, we propose to study whether the information embodied in folksonomies can be used to make network decisions in opportunistic networks.

## 1. INTRODUCTION

The networking community has recently become interested in delay-tolerant or opportunistic networks that aim to transfer data even when there is no complete path between two mobile nodes at any given instant. The idea is to transfer data hop-by-hop, when two nodes meet. Examples of such networks include vehicular networks on highways that rely on predictable car-to-car mobility patterns [7] and "Pocket Switched Networks" which use human mobility [3].

Several basic questions arise in the design of such networks [1, 4, 5]:

1. How do we form "good" topologies during brief interconnection times?
2. How does a node choose the best next-hop node for the rest of the route?
3. How does a resource-constrained node decide whether to accept data for next-hop transfer?

An offline (for instance, trace-driven) analysis can easily compute the answers to such questions. However, finding an online algorithm (if one exists) that works under all operating conditions is challenging.

_____

[*]Currently a visiting student at MIT CSAIL.

## 2. A GENERAL PURPOSE SOLUTION

Clark et al. [2] have proposed a new construct, the Knowledge Plane (KP), which has a global high-level view of the purpose of the network, and can provide advice and other services to elements in the network as needed. The KP is envisioned as using cognitive and AI techniques that are likely to be more general-purpose than traditional, algorithmic approaches which solve individual problems. The knowledge plane could serve as an approximate version of the idealised offline analyser and generically provide online solutions to questions of the sort raised above.

### 2.1 Bootstrapping the Knowledge Plane

Clearly, the KP is data-intensive. Bootstrapping the network with a sufficiently sound and sufficiently complete knowledge base for the KP to reason over is the most significant challenge to realising the promise of the knowledge plane. We approach this problem by making two simplifications.

As originally conceived, the KP is all-knowing, and needs to be able to answer questions such as "Why can't I get to acm.org?". This requires the KP to contain transient information such as whether the user's network interface is up, whether ACM's webserver is currently functional etc. Maintaining a consistent view of such transient information across the network is difficult even in the fully connected Internet and impossible in a delay-tolerant network. We weaken the KP definition by not attempting to store any transient information.

The second simplification is related to knowledge acquisition. Traditionally, knowledge is acquired by planting sensors in the network or from manual input. Sensors can result in abundant but noisy data (i.e. the knowledge base is not sound) whereas user data could be sparse because of inertia to data entry (incomplete knowledge base).

We turn our attention instead to a specific form of knowledge: On Web2.0 sites such as `http://del.icio.us`, users categorise URLs they wish to bookmark with _keywords_ known as "tags". Such tag based sites have quickly become popular, resulting in a huge amount

of publicly available metadata. Collectively, the tags represent a user-defined vocabulary and taxonomy that has come to be known as "folksonomy".

The usefulness of folksonomies comes from being able to relate data based on tags. Two tags are related if some piece of data is tagged with both. The number of co-occurances determines the closeness of their relation. We determine the closeness of data by examining the closeness of their tags. For tags that do not co-occur, we can determine closeness by finding a chain of one or more tags that co-occur and link them together.

Our central hypothesis is that using just closeness of data information, the KP can give intelligent answers to many questions. Note that we only wish to use the folksonomy as the basis for intelligent heuristics rather than generate authoritative answers.

## 3. FOLKSONOMY-BASED REASONING

We illustrate our approach by sketching answers to some of the questions raised in the introduction, within the context of pocket-switched networks [3] where users carry mobile devices that exchange data when the two users come into contact (e.g. in a social setting). Data is forwarded hop-by-hop until it reaches the destination. The devices have finite local storage which is partitioned into a space for the user's own content and a separate space for forwarding content to neighbouring devices.

Now consider Question 3: When should a resource constrained node accept data for next-hop transfer? One heuristic would be to accept data that is closely related to the user's own content or to other content being forwarded. The reasoning is that such data is likely to be useful in satisfying other content requests of the user or to the user's close social contacts. The tacit assumption is that user communities self-select themselves based on the content they are interested in.

Next we turn to Question 2; choosing next-hop nodes. Our strategy is to have the sender forward data to a user with a lot of content that is "close" to the data being forwarded. The hope is that the next-hop user is in the same content-based social sub-community as the receiver. Failing that, the sender can try to forward to a user who has a large number of tags. Such users are likely to be the hubs of the social network and likely have good contacts to find a suitable next hop.

Similarly, to answer Question 1 on forming good topologies, we apply the above concepts to the nodes themselves and preferentially connect closely related nodes.

Notice that the cumulative effect of these heuristics is to convert the difficult problem of guessing the future meeting points of two nodes into the more tractable problem of determining whether the data they intend to transfer are closely related. We emphasise that this substitution is merely a heuristic and may not work in all situations.

## 4. DISCUSSION

It is to be seen whether the success of tagging in the Web2.0 space can be replicated and reused in opportunistic networks. Mathes [6] attributes the success of tagging to the low barrier to entry, and minimal requirements of time, effort and cognitive cost. Our expectation is that if tagging is used to make significant network decisions, users will see a direct benefit to tagging their content and will thus be incentivised to tag. Folksonomies also have the pleasing property that the number of tags and usefulness of the system rises in proportion to the number of users, and thus scale naturally as more users join.

Tag-based systems also have significant drawbacks. Because it is an uncontrolled vocabulary, the tags in a folksonomy can be too personal to have value in shared usage (e.g. a URL tagged "to_read"). Tags can also be ambiguous at times: [6] cites the tag "filtering" being used to describe both water filters and bayesian statistical analysis. Such systems are also unable to automatically see the relatedness of some close but lexicographically distinct tags (Computer_Networks vs. ComputerNetworks or Network vs. Networks). Furthermore, tags are merely string labels, and in their basic form, do not capture numeric data. Two useful examples are the time & date of creation of a piece of content or the number of hops it has traversed. Numeric data can be useful for sorting and in picking the "best" choice.

## 5. CONCLUSION

Using a folksonomic approach, we propose to create a framework for making content-related network decisions by using metadata about the content. Our goal is to design a simplified KP that can be used to answer a wide variety of questions in opportunistic networks.

## 6. REFERENCES

[1] Aruna Balasubramanian, Brian Levine, and Arun Venkataramani. DTN routing as a resource allocation problem. In *Proc. ACM SIGCOMM*, 2007.

[2] D. Clark, C. Partridge, J.C. Ramming, and J.T. Wroclawski. A Knowledge Plane for the Internet. In *Proc. ACM SIGCOMM*, 2003.

[3] P. Hui et al. Pocket Switched Networks and the Consequences of Human Mobility in Conference Environments. In *Proc. SIGCOMM DTN Workshop*, 2005.

[4] Pan Hui and Jon Crowcroft. Bubble rap: Forwarding in small world dtns in ever decreasing circles. Technical Report UCAM-CL-TR-684, Cambridge Univ., Comp. Lab., 2007.

[5] Pan Hui, Eiko Yoneki, Shu yan Chan, and Jon Crowcroft. Distributed community detection in delay tolerant networks. In *Sigcomm Workshop MobiArch '07*.

[6] Adam Mathes. Folksonomies - cooperative classification and communication through shared metadata. Computer Mediated Communication - LIS590CMC, Graduate School of Library and Information Science, University of Illinois Urbana-Champaign, Dec 2004.

[7] Jörg Ott and Dirk Kutscher. A disconnection-tolerant transport for drive-thru internet environments. In *INFOCOM*, pages 1849–1862, 2005.